be allowed to decide what's best for them. Isn't that what drives a free market economy and results in the greatest economic efficiency?

The robustness of this country's computer and software industries is proof that great efficiency, innovation, and productivity can be achieved quickly when industry standards are *voluntarily* set in response to demand. Voluntary standards work. Look at cellular telephones. The FCC recognized that the detailed standards it originally prescribed for cellular telephony were holding back technological development in that industry, and it decided to relax its standards and let the industry establish more advanced standards with minimal government oversight. In doing so, the Commission acknowledged that too much government-specification of industry standards can inhibit technological progress and the availability to consumers of improved services. With Personal Communications Service, or "PCS," the FCC took an even more liberal industry-based approach to standards-setting. It should do the same with digital TV.

Our domestic computer and software industries -- like many other industries -- have thrived in large measure because of two factors: a minimum of government regulation , and open system architecture that permits hardware and software produced by many different firms to interconnect smoothly and encourages rapid, market-driven innovation. Both of these factors would be negated by the FCC's adoption of the Grand Alliance DTV standard, and the public would pay the price.

Let's look for a moment at that standard Beyond public policy and macroeconomic, free-market considerations, there are both consumer interests and technical drawbacks that make adoption of the standard bad policy.

First, the standard does not provide for a way to manufacture low cost receivers. The encoding technique is monolithic If a broadcaster chooses to send the highest resolution format a receiver must include all of the circuitry necessary to decode that format. In a layered system, two signals are sent in the channel simultaneously. A low resolution easily decodable version for smaller cheaper receivers and a higher resolution detail enhancement signal for use by larger, more expensive high definition receivers. In the ATSC system, all receivers, even a little 2" portable must be burdened with means to decode resolution only perceivable on a large screen home theater unit. We have determined that even five years from now a full ATSC decoder will be three times the cost of a base layer decoder. Using the ATSC system will drive up the cost of smaller devices and require consumers to pay for capabilities they may neither need not want.

Second, from a technical perspective, the Grand Alliance standard is a poor compromise, particularly with respect to its video formats. The standard incorporates an obsolete technology, interlaced scanning, that produces an inferior picture and makes inter-conversion for computer uses difficult. In fact, ABC recently announced at a meeting of its affiliates that the network is leaning heavily toward the use of progressive scanning for all its high-definition TV production, because progressive scanning produces a better picture and is less expensive. Even ACATS has admitted that progressive scanning is better. Interlace was an appropriate scheme for the analog television of 40 years ago, but it has no place in a modern digital compressed transmission system.

But broadcasters have been using interlaced scanning for over 40 years. Despite what ABC has said, local stations will have little incentive to replace it with progressive scanning if the FCC adopts a digital standard that allows them to continue to use interlaced. And this is a critical issue for the computer industry because interlaced scanning is unacceptable for text and other computer applications. Any interlaced transmission would have to be converted at the receiver if it is to be used with a computer application. Again, added costs for the consumers.

These limitations of the ATSC proposal would make it more expensive for the domestic computer and software industries to create products -- both hardware and software -- that could enhance the usefulness of digital TVs by marrying digital broadcasting and computers. For these reasons, when ACATS voted to recommend the ATSC standard to the FCC, I abstained.

NTSC broadcast television is transmitted in an analog format. Computer data is digital. As long as analog broadcasting continues, the convergence of TVs and computers will be delayed. But with the advent of digital TV, interactive applications, multimedia, and data sharing between TV and computers are all possible. The products and services that data sharing could make possible are limitless. Microsoft and other firms have committed hundreds of millions of dollars to research and development of products and services that combine computers and TVs; but these products may never reach the stores, at least not at affordable prices, if overly detailed and restrictive regulatory requirements obstruct full compatibility, product development, and competition.

The Grand Alliance says that its proposal provides "adequate" compatibility with computers. We disagree. True, some of the 18 video formats are consistent with computer applications, but the standard also includes a number of inconsistent formats. And if a mandated standard incorporates even one computer-unfriendly format, receiving equipment will need to perform additional conversion and decoding of transmissions to enable interaction with computer applications, the added cost of which will fall on the consumer.

Why does the computer industry care about these issues? Two reasons, mainly. First, we don't want government regulation to freeze technological development without a compelling justification. We think a better DTV standard is possible, and we want the freedom to try it out on the market. Second, our industry knows that computers and TVs can, and will, converge, and we want to be able to develop products that take advantage of that convergence and bring new benefits to the public. Who knows how advanced our National Information Infrastructure can become, if it is allowed to.

In short, in this case, we think voluntary industry standards are better for everyone than government-mandated standards. We understand the value of minimal government-sanctioned technical transmission standards for digital broadcasting, including standards for low level digital bitstream format and modulation technique to prevent interference with other services and would not object to adoption of the ATSC's proposals with respect to those parameters, absent any specified video format.

But specifying a video format is unnecessary and potentially problematic -- exponentially so with 18 formats. We think the marketplace should dictate what video formats it wants. But if the Congress and the FCC find that the public interest would be served by the FCC's adoption of a standard video format for digital television, the standard it adopts should be the best possible. That would not include the hodgepodge of 18 different video formats the FCC is currently considering. If a standard is to be adopted at all, CICATS would propose a simpler, more technologically advanced minimum standard, offering wider compatibility and more flexibility to develop enhancements, if the marketplace warrants.

A year ago, computing capability was not sufficient for the level of convergence of TVs and computers and the sophistication of applications we are developing. It is now. Largely because computer technology is advancing at an exponential rate, the computer industry's interest in advanced television emerged relatively recently. The question should not be *whether* TVs and computers will ever converge seamlessly, but *when* and whether it will be affordable. If the FCC adopts the proposed ATSC standard, the "when" will be years from now -- some say 5 to 7 years later than if the Commission adopts a simpler standard or no standard at all. And when convergence finally arrives, the average consumer will be hard-pressed to afford the advanced products and services convergence will spawn if government regulation imposes a cumbersome, overly complex DTV standard.

If the price of digital receivers and decoders is unnecessarily inflated, the day stations will migrate to all-digital broadcasting will be delayed, and so, in turn, will the day analog spectrum is freed for new uses. In the meantime, precious spectrum could be wasted and consumers could be deprived of better, and cheaper, products and services.

Thank you for your time. I would be pleased to answer any questions you might have.

K

A Video Compression Efficiency Analysis using Progressive and Interlaced Scanning

Eric Petajan

AT&T Bell Laboratories
Murray Hill, NJ 07974

Introduction

The delivery of video programming to the consumer at a reasonable cost and with the highest picture quality depends on a variety of technologies and systems. Individual scenes are transduced with video cameras, film cameras followed by telecine, or reduced by computer. The video signals are then stored on analog video tape or digitized and stored on tape, disk, or electronic image buffer. A finished program is produced by editing individual scenes together. For the last 50 years programs have been delivered to the consumer using the NTSC system. Consumer grade video tape has more recently provided a program delivery alternative to broadcasting. Today we are on the verge of introducing motion compensated video compression into the program delivery process. The consequences of this are far reaching and affect the traditional economics of the entire process. In particular, the choice of video scanning format affects the cost and quality of the video compression to varying degrees depending on scene content. This paper provides an analysis of the relationship between scanning format, scene content, and video compression efficiency as it affects picture quality.

Source Material Preparation

In the interest of conserving computing time and storage, a frame size of 704 H x 480 V was chosen. The 60 frame per second progressive scenes were derived from progressive high definition source material which was appropriately filtered and resampled to 704H x 480V. The interlaced scenes were than derived from the progressive scenes by selecting the odd lines from the odd progressive frames and the even lines from the even progressive frames. Of course, the interlaced scenes have an effective vertical resolution which is significantly lower than the progressive scenes[1].

Video Coder Configuration

A software implementation of an MPEG-2 coder[2] was used with progressive refreshing (see below). No B-frames (bidirectional prediction) were used since the benefit.of B-frames is independent of scanning format. A bit-rate of 4 Megabits/sec was chosen for all experiments, except for the coding of random noise because of its difficulty. The refresh rate was selected to achieve a startup in one third of a second for both formats. Field/frame coding was used for all interlaced scenes. Figure 1 illustrates how the encoder can select whether to construct a given block of pixels from an interlaced frame or from two fields.

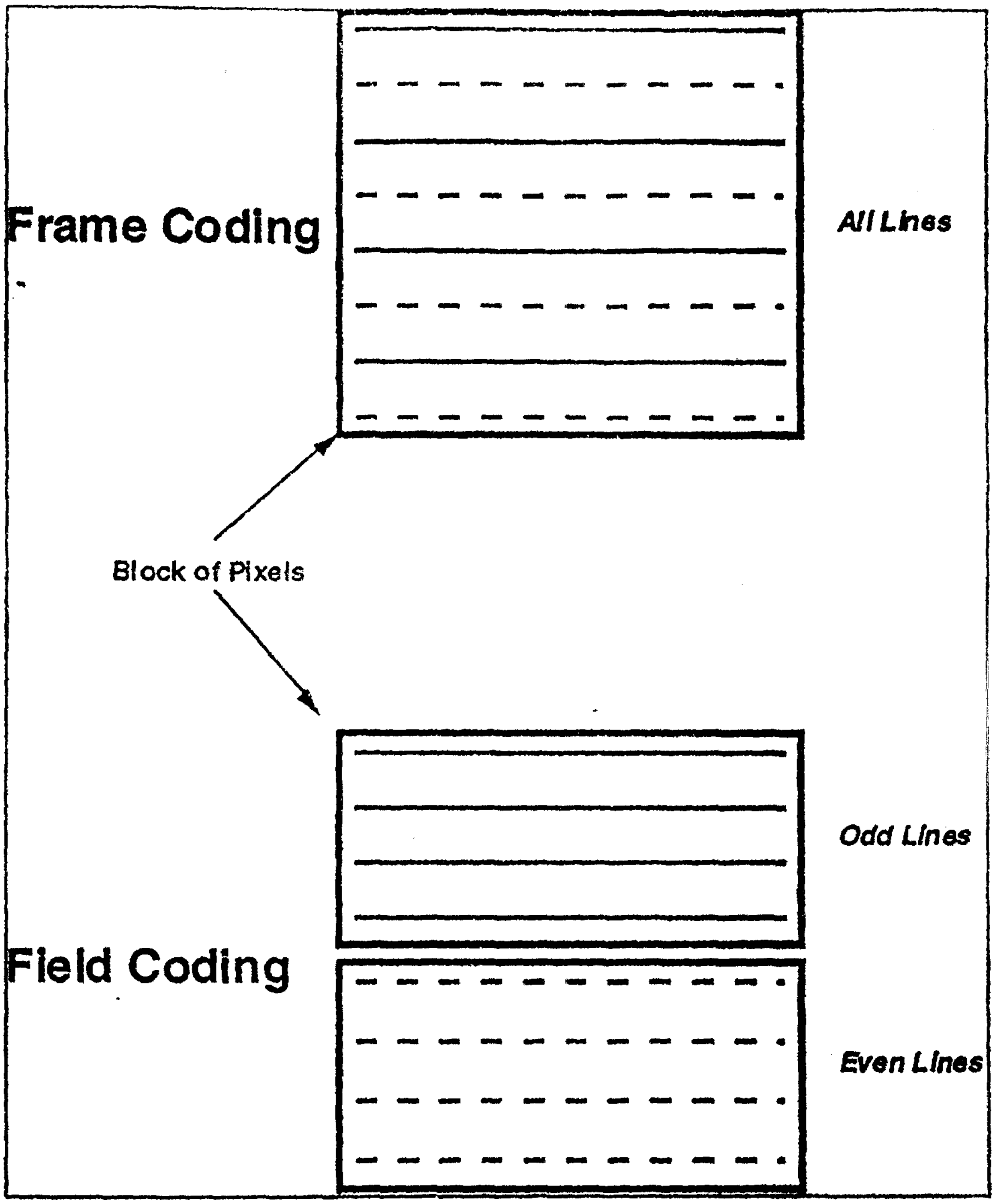

Figure 1. Field/frame coding

The picture quality was measured using the mean squared error of the difference between the coded and the original pictures. This was expressed as a signal to noise ratio in decibels using the following equation:

$$SNR = 10 \log_{10} [255^2/(MSE(\text{coded picture})]$$

It is generally accepted that differences in SNR of less than .5 dB are not significant.

Static and Predictable Scenes

Motion compensated transform coding explicitly measures spatial and temporal redundancy in an image sequence and only sends unique picture information to the decoder (see Figure 2). The use of intra-frame-only coding (refreshing shown in Figure 3) for decoder startup (channel acquisition), or to provide insert edit points, is an exception to temporal redundancy removal in the encoding process and requires an increase in coded bit-rate to maintain equivalent picture quality. The best illustration of this is in the coding of a static image sequence (repeated still). Virtually the only information required by the decoder after startup is a set of zero-length motion vectors for each frame which consumes a tiny fraction of the bit-rate for a motion sequence. However, the use of I-frames or I-blocks (I means intra-frame coding) dramatically increases the bit-rate to levels comparable to coded motion scenes.

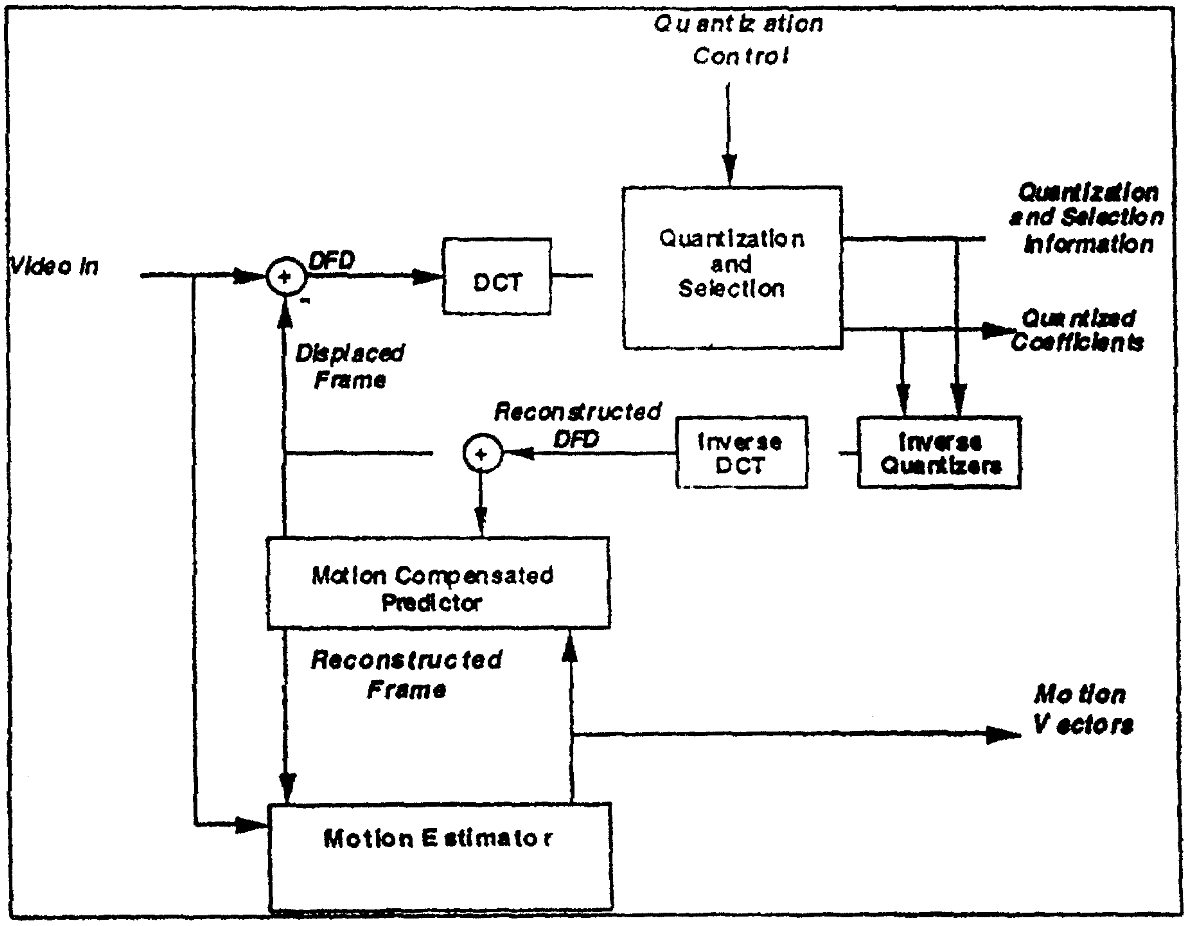

Figure 2. Video Encoder Loop

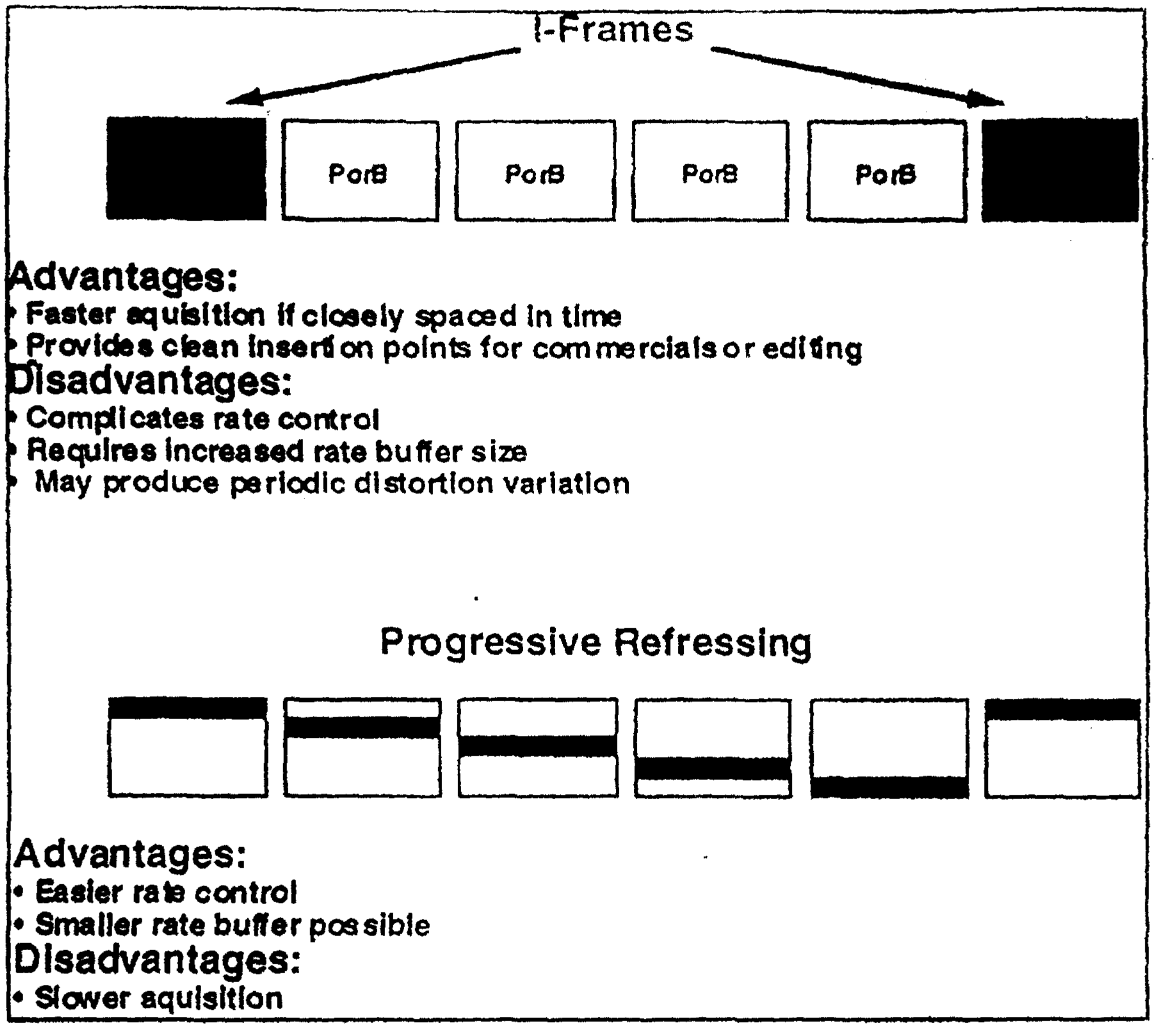

Figure 3. Refreshing techniques

To achieve a given decoder startup time or insert edit point period, an entire frame must be intra-frame coded within the given time constraint. Since the frame rate in our progressive format (60 frames/sec) is twice that of the interlaced format (30 frames/sec), the ratio of intra-code frames to inter-coded frames must be twice as high for the interlaced format compared to the progressive format to achieve the same decoder startup time. Therefore, the number of intra-coded frames per second is equivalent between our interlaced and progressive formats. This holds true for both I-frames and progressive refreshing with I-blocks. Since virtually all of the bit-rate from a coded static scene is consumed by intra-frame information, the coded picture quality should not depend on whether interlace or progressive scanning is used. However, the coding process will not remove interlace artifacts. Thus, for static scenes, progressive scanning provides equivalent coded picture quality compared to interlaced scanning without interlace artifacts. This was verified experimentally and the results are shown in the first row of Table 1. The image of Chicago was coded with an SNR of 39.83 dB using progressive and 39.97 dB using interlaced scanning. This .14 dB difference is not significant.

Scene	Bit-rate	Progressive SNR(dB)	Interlaced SNR(dB)	Prog SNR - Int SNR
Chicago Still	4 Mbits/sec	39.83	39.97	-0.14
Panned Map	4 Mbits/sec	21.92	21.84	0.08
Noise	12 Mbits/sec	18.10	19.57	-1.47
Chicago Zoom	4 Mbits/sec	27.19	26.91	0.28
Mall	4 Mbits/sec	34.61	34.96	-0.35
Traffic	4 Mbits/sec	39.40	38.58	0.82

Table 1. Video coding results

The second row of Table 1 shows results for a Panned Map which is highly predictable and contains no noise. As expected, the two formats performed nearly equally with the progressive SNR higher than the interlaced SNR by .08 dB.

Random Noise

Now consider the coding of a sequence of frames of random noise. This type of scene is the opposite of a static scene from a video coding perspective, i.e., static scenes are completely correlated (at least temporally) and noise is completely uncorrelated. The only opportunity for redundancy removal in this case is the substitution of coding artifacts for some of the random noise using human perceptual modeling. Again, the intra-coded block rate is equivalent between our two formats but now the inter-coded blocks consume nearly as many bits as the intra-coded blocks and the interlaced format has half as many inter-coded blocks per second as the progressive format. Therefore, the coding of interlaced random noise should provide better fidelity than progressive random noise. In effect, interlaced scanning of random noise discards half of the noise samples before coding which reduces the bit-rate proportionately. The third row of Table 1 shows the experimental results for this case where the coding of a noise sequence produced a 1.5 dB increase in SNR using interlace compared to progressive scanning. A bit-rate of 12 Megabits/sec was used for this difficult scene to give reasonable SNR values.

Typical Scenes

Row 4 of Table 1 shows coding results for a scene which contains no noise but is only partially predictable because it is a computer generated zoom using the Chicago still. Block-based motion compensation can only approximate non-translational motion such as zooming or rotation. Progressive scanning is slightly favored for this scene with a .28 dB increase in SNR compared to interlace.

Typical camera scenes contain some noise (electronic or film grain), static or temporally predictable areas (panning), and areas with unpredictable or complex motion (uncovered background, fast zooms). The contribution to the total coded bit-rate from each type of scene content is proportional to the area of each type integrated over the duration of the scene. The contribution to coded bit-rate from noise is proportional to the noise amplitude and spectral characteristics. Table 1 lists two scenes in rows 5 and 6 which were filmed at 30 frames/second called Mall and Traffic. These scenes were scanned and digitized before coding and they were doubled in speed to 60 frames per second in order to derive both 60 frames/sec progressive and 30 frames/sec interlace from the same scenes. Of course changing the frame rate in simulation is done merely by changing a software parameter. The Mall scene was shot indoors and contains the random motion of a fountain and some complex motion (people walking). Increased film grain from indoor light levels and random motion gives the interlaced form of this scene a .35 dB increase in SNR compared to the progressive form. This is not significant and does not result in any visible improvement in picture quality. The Traffic scene was shot outdoors and contains various speeds of motion. The progressive form of this scene produced a .82 dB increase in SNR compared to the interlaced form. This is a somewhat visible difference in picture quality. The interlaced forms of both scenes contain visible interlace artifacts.

Conclusions

The experimental results clearly show on a wide variety of scenes that the picture quality of coded progressive scenes is equal or better than that of the interlaced form of the same scenes. In one case the progressive picture quality was significantly better than interlaced (not considering interlace artifacts). This may have been due to the increase in spatial frequency energy in moving areas. If frame coding is used, moving edges are jagged leading to high frequency DCT coefficient amplitude. If field coding is used, the smaller block size reduces the efficiency of the DCT.

Since the pixel rate of the progressive format is twice that of the interlaced format, the coding efficiency for progressive scanning has been shown to be twice that of interlaced scanning. The only exception to this is scenes with high amplitude random noise. Properly coding such scenes calls for noise filtering before coding using progressive scanning. If the noise was intentionally added for effect then a block-based pseudo-random noise pattern should provide sufficient spatial and temporal redundancy for good picture quality. If the availability of progressive scan cameras is in question then deinterlacing before video coding should provide most of the benefit of progressive scanning.

References

1. C. Vogt, "Camera scanning standards in future television systems", EBU Review No. 244, December 1990.

2. "ISO CD 11172-2: Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s," Nov. 1991.

L

Progressive versus Interlaced Coding

Philippe Guillotel† and Stephane Pigeon*

†THOMSON MULTIMEDIA R&D FRANCE
1, Avenue de Belle Fontaine
35510 Cesson-Sévigné, FRANCE
Tel : +33 99.25.42.00
Fax : +33 99.25.43.34
E-Mail : guillotelp@tcetbs1.thomson.fr

*UNIVERSITE CATHOLIQUE DE LOUVAIN
2, Place du Levant
B-1348 Louvain-La-Neuve, BELGIUM
Tel : +32.10.47.23.12
Fax : +32.10.47.20.89
E-Mail : Pigeon@tele.ucl.ac.be

Abstract

Interlaced versus progressive scanning is an important issue when dealing with digital television. Not only because the change from analog to digital communication may be seen as an opportunity to move to other formats, but also because of the well-known artifacts of interlaced scanning (interline twitter, line crawling, and field aliasing) compared to the natural way of representing two-dimensional images as the progressive format does. However, digital broadcasting has to face the problem of transmitting twice the number of pels of the progressive format. It is the purpose of this article to study this problem, and especially to check if the increased vertical and temporal correlations of the progressive pictures provide a significant improvement in the bit-rate reduction efficiency. In that case, progressive scanning may also be used as an intermediate transmission format to improve the compression performances of interlaced sequences.

1. Introduction

Interlaced scanning was introduced about 25 years ago as a simple and effective trick to halve the bandwidth, resulting in a shape size in the vertical/temporal domain adapted to the human vision limitations, hence its high spatial definition and field rate. However, critical material emphasizes typical interlaced artifacts, such as the well-known interline twitter, line crawling and field aliasing[1]. These defects are much more annoying today because of the improved picture quality of both displays and cameras. Moreover, half the bandwidth for analog transmission of TV signals is an efficient solution, whereas for digital communication the challenge lies in achieving a high picture quality at a given bit rate. This requirement in the coding efficiency leads to the MPEG-2 standard [2].
From these considerations progressive scanning can be considered as a candidate for a new transmission format, because progressive pictures have higher vertical resolution, seem much more attractive than interlace for signal processing, and guarantee the compatibility with other multimedia applications. Unfortunately, the number of samples is twice that of the existing interlaced format.

It is the purpose of this paper to compare the efficiency of both progressive and interlaced formats in the context of a MPEG-2 coding scheme. Based on these results different conclusions will be drawn to demonstrate that the progressive format improves the overall picture quality, and that such a transmission format may be also an intermediate step towards progressive broadcasting without loss of performances compared to the existing interlaced one. Unfortunately the compression performances can not be significantly increased.

2. Coding Efficiency Comparisons

The included simulation results are obtained from two different MPEG-2 broadcasting chains in both scanning formats (details in [3]), and with the following source materials (results for the four last progressive sequences are available only with interlaced display) :

• **Interlaced :** *Mobile and Calendar* and *Flower and Garden* : From a tube camera;

• **Progressive :**

# *Renata RAI*	: From an HDTV tube camera;
# *Kiel Harbor* and *Kiel Harbor 2*	: Digitized photo with synthetic motion;
# *Pendel* and *Foot*	: From a progressive tube camera;
# *Pops*	: From a progressive CCD camera;

Two different deinterlacers, one at the transmitter side (high quality motion compensated [4]), one at the receiver side (low cost macroblock based solution, making use of the transmitted MPEG-2 motion vectors), deal with the interlaced to progressive conversions (more details can be found in [5]). The opposite format changes are performed through vertical filtering (including the Kell factor) and subsampling.

In addition, two bit-rates have been selected (4 Mbit/s excepted *MOBILE* encoded at 6 Mbit/s) in order that the picture quality over all the set of sequences is constant (PSNR between 30 and 35 dB). The PSNR (Peak Signal to Noise Ratio) together with a subjective expert analysis evaluate the efficiency of each scenario.

2.1 MPEG-2 Encoding Parameters

Some parameters have to be defined to comply with the MPEG-2 syntax. Among them some are specific to the progressive format and can be optimized such as :

- *progressive_frame* set to 1, coded video contains only progressive frame pictures. It leads to : *picture_structure*= "frame" and *frame_pred_frame_dct*=1;
- *frame_pred_frame_dct* set to 1. For each macroblock, this flag suppresses useless flags like *frame_motion_type* (2 bits) and *dct_type* (1 bit) from the bitstream;
- The motion estimator is a 5 hierarchical levels block-matching with a [-127,+128]×[-63,+64] half-pel vector range. It is based on a pyramidal structure which leads to a very simplified and efficient data processing when dealing with progressive (1 vector instead of 5). Furthermore, it leads to a simplified mode decision processor.

Accordingly, progressive coding reduces the side-information by 3 bits/macroblock, it lowers the number of vectors to transmit, and simplifies the chrominance filters.

Other MPEG-2 parameters are identical for both formats such as the VLC intra tables (*intra_vlc_format=1*), the non-intra quantization matrix (flat), the macroblock mode selection, the thresholding of the DCT coefficients, the quantizer type (*q_scale_type*=0), the zig-zag matrix (*alternate_scan*=0). All these points are not in the scope of this paper and will not be further discussed.

The encoder is thus MPEG-2 compliant except for its use of the progressive (not currently supported by this profile). Anyway, the objective of this study is to compare both formats with the same picture size, and a new level might be further included in the MPEG-2 final standard specification to comply with progressive scanning.

Finally, only the GOP structure remains to be specified. For interlaced signals the classical one is used (M=3, N=12) when for progressive pictures computer simulations lead to M=5, N=25 (slightly more efficient than M=6, N=24).

2.2 PSNR and subjective picture evaluation

Let us just remind that between pictures of the same format a better PSNR value generally means a better picture quality if the gap is significant (greater than 0.5 dB), otherwise subjective picture evaluation is required. For instance with the previous display formats, and considering that progressive display leads to a higher picture quality, a lower progressive PSNR value does not necessarily mean a lower picture quality.

- **Interlaced display** (progressive coding + receiver interlacing / interlaced coding + display) :

	Mobile		**Flower**		**Kiel**		**Renata**	
Coding Format	Prog	Int	Prog	Int	Prog	Int	Prog	Int
PSNR (dB) Y	29.32	32.30	30.38	30.64	32.11	31.61	33.49	33.14
PSNR (dB) U	33.90	34.45	33.47	33.39	39.08	39.23	36.07	35.69
PSNR (dB) V	31.85	32.11	31.87	31.38	37.82	38.00	37.86	37.67
	Foot		**Kiel 2**		**Pendel**		**Pops**	
Coding Format	Prog	Int	Prog	Int	Prog	Int	Prog	Int
PSNR (dB) Y	32.23	30.84	29.17	27.81	41.25	41.87	36.35	36.99

Table 1 - PSNR (dB) for interlaced signals

Progressive coding performs slightly better (PSNR and picture quality) for 4 sequences (*Kiel, Renata, Foot, Kiel 2*). For two (*Flower* and *Pendel*) the visual quality is in favor of the progressive format, confirming that the PSNR difference is too low to be significant (*Flower* < 0.3 dB), or too high for visual artifacts (*Pendel*). And finally, *Pops* leads to visually similar pictures (difference = 0.6 dB), and *Mobile* performs better when interlaced coded (+ 1 dB).
Thus the two formats perform similarly (average PSNR : 0.17 dB more for progressive), except when the deinterlacing failed. In addition, the Kell filter, for progressive to interlaced conversion, acts as a post-filter to improve the picture quality of the interlaced decoder.

- **Progressive display** (progressive coding + display / interlaced coding + receiver deinterlacing) :

	Mobile		**Flower**		**Kiel**		**Renata**	
Coding Format	Prog	Int	Prog	Int	Prog	Int	Prog	Int
PSNR (dB) Y	31.30	27.51	31.41	26.59	30.36	26.10	31.12	27.18
PSNR (dB) U	34.26	33.28	34.10	33.68	40.47	39.21	35.55	34.24
PSNR (dB) V	32.29	31.44	32.30	30.83	39.15	37.85	37.47	36.32

Table 2 - PSNR (dB) for progressive signals

The only conclusion from the previous table is that the macroblock based deinterlacer does not perform very well. It means that very simple and low cost solutions can not be used, and that careful design should be done to reach an acceptable quality.

- **Interlaced / Progressive chain** (progressive coding + display / interlaced coding + display) :

	Mobile		**Flower**		**Kiel**		**Renata**	
Coding Format	Prog	Int	Prog	Int	Prog	Int	Prog	Int
PSNR (dB) Y	31.30	32.30	31.41	30.64	30.36	31.61	31.12	33.14
PSNR (dB) U	34.26	34.45	34.10	33.39	40.47	39.23	35.55	35.69
PSNR (dB) V	32.29	32.11	32.30	31.38	39.15	38.00	37.47	37.67

Table 3 - PSNR (dB) for progressive and interlaced broadcasting

From table 3 interlaced broadcasting seems better than progressive except for *Flower*. As a matter of fact, subjective evaluation show that, besides nearly 1 dB loss (for *Mobile*), the picture quality is higher with progressive encoding of interlaced sources because it removes the interlaced artifacts (flicker). In addition, the double resolution of the progressive original pictures explains the PSNR loss when progressive encoded, but the picture quality can be higher (fixed and detailed areas of *Kiel*), or lower (interlaced effects sometimes masks the coding artifacts of *Renata*) depending on the scene content.

From the three previous analysis, the following conclusions can be pointed out :

1)- An all progressive chain is generally preferred to an all interlaced one;

2)- In case of interlaced display, progressive transmission improves the picture quality of progressive sources compared to their interlaced versions, and the loss of resolution with interlaced sources (due to the interlacing filter) can supersede the reduction of blocking effects brought by the progressive encoding.

To explain these results, the following classification has to be done between sequences with similar vertical resolution and sequences with different vertical resolution, but also depending on the motion content. It leads to table 4.

1)- Without motion (*Mobile, Pendel, Pops, end of Renata*) : The pictures are frame coded in both formats, thus the spatial correlations and the motion performances are similar. The double number of pels of the progressive leads to a double bit-rate for I frames, but also for B frames since twice the number of vectors have to be transmitted (the bit-rate required for the macroblock header including motion vectors is 30% to 40% of the total bit-rate). For P frames the motion estimator performs better with progressive scanning (lower temporal distance), and the bit-rate required for the macroblock header represents less than 20%. However, it is not enough to prevent the 1 dB loss moving to progressive scanning in the case of interlaced source pictures, and this is increased up to 3 dB loss for progressive sources pictures because of the increased resolution;

2)- With motion (*Flower, Foot, Kiel, Kiel 2, beginning of Renata*) :The pictures are field coded. The number of motion vectors is the same in both case (2 fields vectors are transmitted per macroblock). It can thus be expected to have the same bit-rate for the B frames whatever the format is. In addition, once again progressive performs slightly better for the motion prediction, the bit-rate is thus expected to be lower than twice that of the interlaced P frames. Finally, the spatial correlation is probably better for progressive pictures, the bit-rate for I frames should not be too much higher than in the interlaced case. The result is 1 dB gain moving towards progressive scanning with interlaced source signals and 1 dB loss with progressive source signals (once again the additional 2 dB loss is due to the increased vertical resolution);

	Interlaced source		**Progressive source**	
	Static	Motion	Static	Motion
Prog/Int coding + Int display (Int/Int PSNR)	-3	-1	-1	+1
Prog/Int coding + display (Prog/Int PSNR)	-1	+1	-3	-1

Table 4 - PSNR gain (dB) moving towards progressive scanning

When interlaced display is performed for each format, 2 dB have to be subtracted to the performances of the interlaced original pictures, and 2 dB have to be added to those of the progressive sources (the first gain is due to an average value computed with less samples, and the second loss to a filtering effect).

2.3 Influence of the Bit-Rate

Is the comparison between progressive and interlaced scanning bit-rate dependent ? To answer this question, simulations on the sequence *Pops* have been performed at 2, 4 and 6 Mbit/s considering interlaced display. Table 5, clearly shows that if interlace is better at high bit-rates this is still true at low ones if not even more (from 0.6 dB at 6 Mbit/s, up to 1.7 dB at 2 Mbit/s). The number of pels as well as the vertical and horizontal resolution are very critical at low bit-rates, and, even with interlacing, prefiltering is often required to smooth the picture content. If at high bit-rates the increased vertical resolution can be compensated, it is not true at low ones. Consequently, the performances of the progressive format decrease faster than those of the interlaced one at low bit-rates.

Bit-rates	**2 Mbit/s**		**4 Mbit/s**		**6 Mbit/s**	
Coding Format	Prog	Int	Prog	Int	Prog	Int
PSNR (dB) Y	32.17	33.87	36.35	36.99	37.98	38.58

Table 5 - PSNR (dB) at different bit-rates

2.4 Influence of the Picture Complexity

It has been shown that the conclusions differ depending on the picture content. Table 6 sum up the previous results by decreasing order of complexity value, referring to the original progressive sequences that have been interlaced. The PSNR can be considered related to the difficulty to encode a picture, thus it is selected as complexity measure (a high complexity gives a low value)

	Kiel 2 (28dB)		**Foot (31dB)**		**Kiel (32dB)**		**Renata (33dB)**		**Pops (36dB)**		**Pendel (41dB)**	
Coding Format	Prog	Int	Prog	Int	Prog	Int	Prog	Int	Prog	Int	Prog	Int
PSNR (dB) Y	29.17	27.81	32.23	30.84	32.11	31.61	33.49	33.14	36.35	36.99	41.25	41.87

Table 6 - PSNR (dB) for different picture complexity

From table 6, progressive performs clearly better for complex images and a little worse for pictures with a low complexity. The reason is that at low complexity the progressive format bring no additional information compared to interlace, and since twice the number of lines should be transmitted it results in slightly lowering the PSNR of the decoded pictures. However, since the gap is nearly equal to 0.5 dB, and since both progressive and interlaced PSNR are high, no noticeable difference between both formats can be seen.

2.5 Influence of the Deinterlacing

Moving towards progressive transmission will require conversions from progressive to interlaced and interlaced to progressive scanning to manage present studio environment. Thus the effects of the deinterlacing have to be studied to be sure that it handles field aliasing properly. Table 7 depicts the results of simulations performed on the Kiel 2 progressive source sequence by means of PSNR values (they refers to the original sequence that has been interlaced allowing for reliable comparisons). The original pictures are progressive encoded and interlaced displayed to give the PSNR value called *progressive* in table 6. Then the source is interlaced coded and displayed, and its PSNR computed in column *interlaced*. Finally, the previous interlaced sequence is *deinterlaced* to go back to progressive coding and final interlaced display.

Coding Format	Progressive	Interlaced	Deinterlaced
PSNR (dB) Y	29.17	27.81	28.36

Table 7 - PSNR (dB) between interlaced, deinterlaced and progressive signals

As expected, the deinterlaced sequence is better than the interlaced one, because the original progressive source performs already better than the interlaced version, and because the deinterlacing is artifacts free on that sequence.
However, these results are very dependent on the quality of the deinterlacer, thus conclusions may take into account possible low quality deinterlacing. Having in mind that future deinterlacing will become better and better.

3. Conclusion

In this paper, the coding efficiency of both progressive and interlaced scanning formats are compared by means of PSNR values and subjective picture quality analysis. The main goal was to evaluate the impact of using a progressive transmission format compared to the existing interlaced one. It leads to the conclusion that the absence of interlaced artifacts (mainly line flicker) allows the use of a greater compression factor in the case of progressive processing and display. At the same bit-rate an all progressive broadcasting chain, from the source capture to the final display, is thus preferable to an all interlaced one, except for an increased hardware complexity if twice the number of pels is scanned. Moreover, with interlaced display, the progressive transmission can be considered at least as good as the interlaced one and generally better if progressive sources are encoded. Unfortunately, the conclusions are not so clear when dealing with interlaced sources : the loss of resolution supersedes sometimes the reduction of blocking effects and the conversion from progressive to interlaced scanning after decoding can either improve (post-filtering of the coding artifacts) or decrease (loss of resolution) the picture quality depending on the source sequences available.
Consequently, it has been shown that progressive does not lead to a loss of performances, that on the contrary it brings a more stable picture quality, even if the MPEG-2 standard has been optimized for interlaced signals.
Thus, from a picture quality point of view, progressive scanning is a very attractive format for the transmission, and even more for the visualization of pictures. In addition, progressive can be used as an intermediate step towards progressive broadcasting of TV signals without loss of performances compared to the existing interlaced format. This is even more interesting when a smaller picture size is considered, to comply with the actual MP@ML profile (of course comparable picture quality is assumed).
Finally, if the MPEG-2 compression performances can not be significantly increased moving towards progressive scanning, compatibility with the multimedia applications (Computer, Broadcasting, Transmission, Virtuality, Film, ...) will be simplified and more efficient. This is perhaps the best way to go to.

Acknowledgment

This work has been supported by the European project RACE 2110 "HAMLET".

References

[1] S. Pigeon and P. Guillotel, "Advantages and Drawbacks of Interlaced and Progressive Scanning Formats", *CEC HAMLET Deliverable N° R2110/WP2/DS/R/004/b1*, Y2/M6 1995.

[2] International Organisation for Standardisation, "Generic Coding of Moving Pictures and Associated Audio", Draft recommendation H.262, ISO/IEC 13812-2, March 1994.

[3] P. Guillotel and S. Pigeon, "Coding Efficiency Comparisons of both Interlaced and Progressive Scanning Formats", *CEC HAMLET Deliverable N° R2110/WP2/DS/R/012/b1*, Y2/M12 1995.

[4] L. Vanderdorpe et al., "Motion Compensated Conversion from Interlaced to Progressive Formats", *Signal Processing : Image Communication*, Vol. 6, No. 3, June 1994, pp. 193-211.

[5] S. Pigeon and al, "Specification of a Generic Format Converter", *CEC HAMLET Deliverable N° R2110/WP2/DS/S/006/b1*, Y2/M9 1995.

A comparative study of simulcast and hierarchical coding

Participants in RACE project R2110 "HAMLET", subgroup WG2 ;
J. De Lameillieure*, S. Pallavicini†

February 1996

Abstract

Simulcast or simultaneous broadcasting of a program at two different quality or resolution levels requires a less complex source coding than scalable or hierarchical source coding, where there is a link between base and enhancement layer. In this paper, we will investigate the conditions under which a scalable system has a better subjective quality compared to simulcast at equal bit-rates.

1 Introduction

Hierarchical coding represents data in two layers : a base layer and an enhancement layer. A hierarchical decoder has to decode both layers and has therefore a higher complexity than a decoder that decodes just one layer. A transmission with no hierarchical coding link between both layers broadcasts simultaneously the same program at two different bit-rates or quality levels, and is therefore called "simulcast". This simulcast scenario does not require a more complex decoder for the enhancement layer. This contribution discusses pros and cons of simulcast and hierarchical source coding over a hierarchical transmission chain.

In section 2 we will first define the hierarchical transmission chain. Section 3 and 4 treat the cases of quality scalability and resolution scalability. Section 5 discusses the hardware complexity of the scalable and the simulcast decoder. Section 6 is a report of subjective tests of spatial scalability and simulcast carried out in the Eureka-project "ADTT".

2 Hierarchical transmission chain

Both hierarchical or scalable coding and simulcast are ways of source coding for a hierarchical transmission chain, i.e., a transmission chain with a channel coding and a modulation that have two levels of protection : a well-protected part that can be received under good and under severe transmission conditions, and a less protected part that can be received under good conditions only. The recently decided specification of DVB for terrestrial transmission [1] foresees such a hierarchical transmission chain as an option. An example of a hierarchical or rugged transmission chain has been demonstrated by the $^{\mathrm{H}}\mathrm{DTV}_{\mathrm{T}}$ project during the IFA exhibition in Berlin, Sept. '95 [2].

In hierarchical transmission chains, the available net bit-rate for source coding is usually smaller for the well-protected part (base layer) than for the less-protected part (enhancement layer). There are mainly two reasons for that :

A first reason is that the higher protection of the base layer by the channel coding requires a proportionally higher gross bit-rate. In practical systems, the gross bit-rate of the base layer does not exceed the gross bit-rate of the enhancement layer. Because of the proportionally higher channel coding bit-rate for the base layer, the net bit-rate for source coding of the base layer is much less than the net bit-rate for the enhancement layer.

Another reason for the lower net bit-rate for the base layer is the hierarchical modulation : a modulation constellation can be configured more robustly at the cost of available gross bit-rate capacity. If base

*Heinrich-Hertz-Institut Berlin, Germany
†CCETT, Rennes, France

and enhancement layer have a comparable part of the energy or bandwidth of the modulated signal, the hierarchical modulation will have a lower gross bit-rate for the base layer than for the enhancement layer. The higher channel coding protection will in its turn even more reduce the available net bit-rate in the base layer.

3 Hierarchy of quality levels

In this section, we will only consider systems where the output of base and enhancement layer have the same spatio-temporal resolution but a different quality of reconstruction. This means for the scalable source coding "SNR scalability" or "Data partitioning" [3, 4].

Similar to the scalable case, we will call the simulcasted bitstreams with lower and with higher bit-rate the "base layer" and "enhancement layer" respectively. The enhancement layer in both simulcast and scalable coding has been compressed with a finer quantisation.

3.1 Picture quality in simulcast and scalable coding

Picture quality can be measured objectively by means of the Signal-to-Noise-Ratio (SNR) or by means of subjective assessments (e.g. according to ITU-R Rec. 500 [5]). Although the measurement of the subjective quality is quite cumbersome compared to the calculation of SNR values, it is the subjective quality that counts in the comparison of different source coding alternatives. In the following, we mean subjective quality when writing "quality".

The picture quality after compression and decompression of a digital video sequence is usually an ascending function of the bit-rate. The quality as a function of the bit-rate is usually steeply ascending for low bit-rates and saturates at higher bit-rates (see Fig. 1)

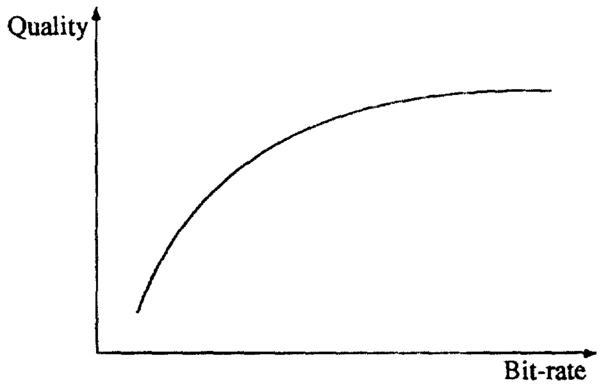

Figure 1: Picture quality vs. bit-rate ; the bit-rate around which the saturates depends on the video sequence.

We will now compare the scalable and simulcast coding. In both alternatives, the base layer has a smaller video bit-rate than the enhancement layer.

3.1.1 Simulcast

The achievable qualities in each of the layers of simulcast is shown in Fig. 2.

Only if the bit-rate of the base layer is sufficiently less than the saturation bit-rate, there will be a visible difference between both layers. Only in that case, a hierarchical transmission chain combined with simulcast makes sense. Otherwise there is hardly any noticeable quality jump between both levels of quality in simulcast.

3.1.2 Scalable coding

In a first approximation, the quality of the scalable enhancement layer corresponds to the quality of the summed bit-rates of base and enhancement layer. The situation is then as depicted in Fig. 3.

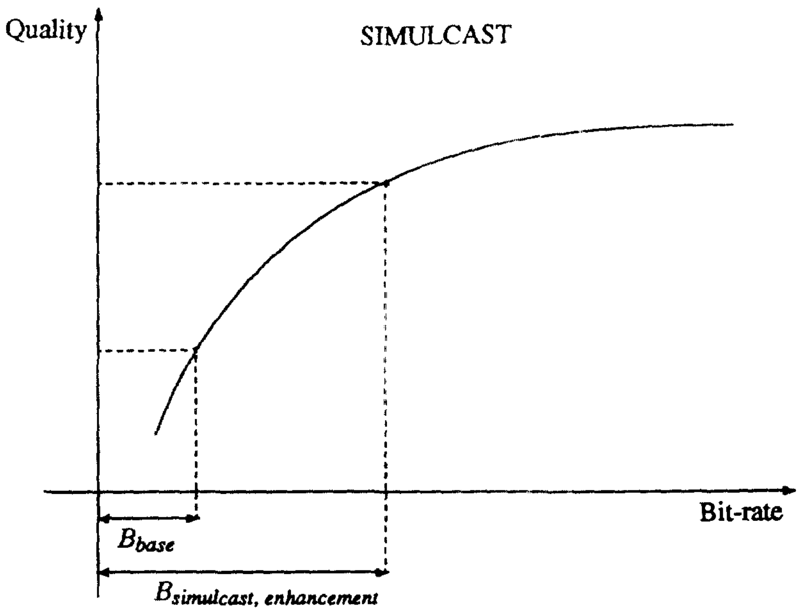

Figure 2: The quality of both layers in a simulcast system

This figure clearly shows that a scalable system has an advantage compared to simulcast when the bit-rate of the simulcast enhancement layer is not higher than the bit-rate where the quality saturates.

The quality of a scalable system is not exactly equal to the quality corresponding to the sum of the bit-rates of base and enhancement layer. The scalability costs some bit-rate for overhead information. As the quality of the enhancement layer is usually near the saturated part of the quality curve, the subjective cost of the scalability overhead is mostly small. This has been confirmed by subjective tests [6, 7], where the quality of SNR scalability with 3+4 Mbit/s was comparable to the quality of non-scalable coding at 7 Mbit/s.

3.2 Discussion

The question on the sense and nonsense of scalable coding in a hierarchical transmission chain (hierarchical channel coding and modulation) is according to the above description completely dependent on the sequence and the available bit-rates in base and enhancement layer. The answer depends on whether the bit-rates are in the range where the subjective quality saturates or not. The answer can be summarized as follows:

1. Simulcast or scalable source coding make only sense in a hierarchical transmission chain when the net video bit-rate of the base layer is sufficiently below the bit-rate where subjective quality saturates.

2. Scalable source coding outperforms simulcast if the net video bit-rate of the simulcast enhancement layer is below the bit-rate where the subjective quality saturates.

In applications where the bit-rates of base and enhancement layer are not variable, it is quite probable that scalability is only advantageous in critical sequences with a rather high quality saturation bit-rate.

4 Hierarchy of resolutions

In this case, the resolution (picture size in pels and/or frame rate) of the enhancement layer is higher than the resolution of the base layer. Each layer can be coded independently (simulcast). Alternatively, the enhancement layer can be predicted by upconversion of the base layer (spatial or temporal scalability).

In the application envisaged by the HAMLET hardware, the base layer is TV while the enhancement layer is HDTV. One could think of simulcast or spatially scalable transmission *without* hierarchical channel coding and modulation just to provide the same program content to low-cost receivers with the lower

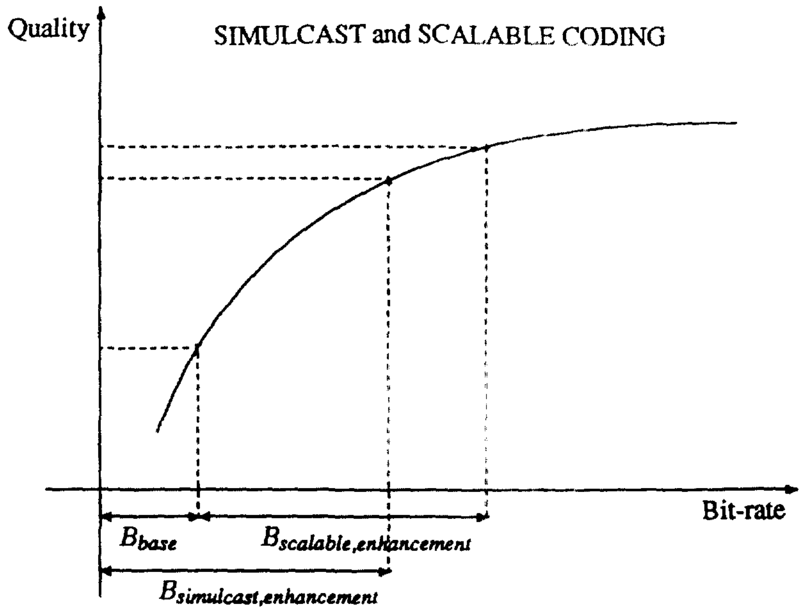

Figure 3: The quality of both layers in a scalable system

resolution and to high-resolution receivers. When the two resolution layers are transmitted *with* hierarchical modulation and channel coding, a graceful degradation in the high-resolution decoder can be realised under bad reception conditions by falling back to the upconverted lower resolution layer. On top of that, the stronger base layer signal will allow the plug-free and portable reception of the broadcasted program, albeit in base layer resolution.

4.1 Picture quality in simulcast and scalable coding

In the case of resolution hierarchy, the same considerations as in subsection 3.1 on subjective quality and saturation bit-rate apply to the upconverted base layer and the enhancement layer.

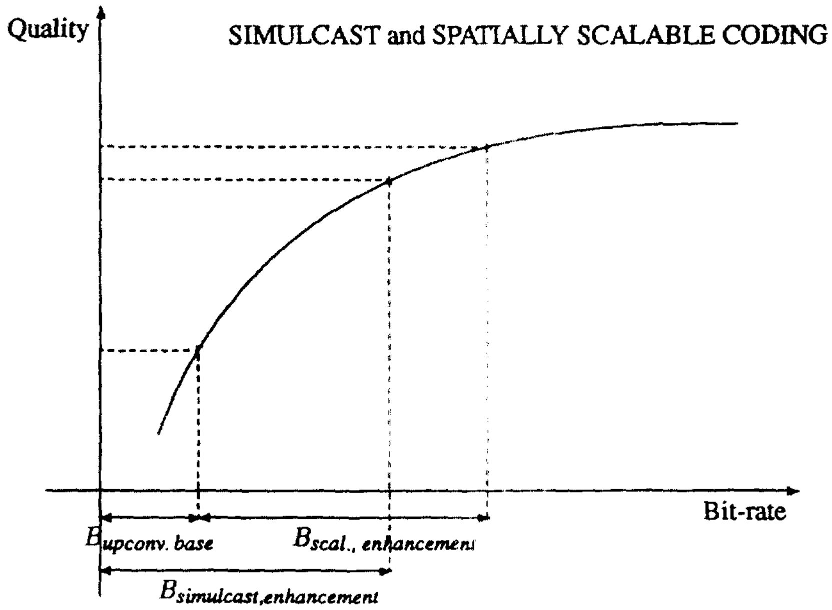

Figure 4: The quality of the upconverted base layer and the enhancement layer in a spatially scalable system

4.2 Discussion

It is an a-priori-choice to include a base layer with lower spatio-temporal quality in the complete system here. As a consequence, there is no conclusion that is directly equivalent to the conclusion 1 of section 3.2. It is only possible to state whether a fall-back to the upconverted base a layer makes sense :

1. A fall-back to the upconverted base layer makes only sense if the upconverted base layer has a quality sufficiently below the saturation quality.

According to our experience with spatial scalability, this is usually the case. Similar to conclusion 2 of section 3.2, we have :

2. Spatial scalability can only have a better quality of its enhancement layer compared to simulcast if and only if the bit-rate of the enhancement layer is below the bit-rate where the subjective quality saturates.

Therefore, spatial scalability can only be advantageous if for a given bit-rate of the enhancement layer the simulcast of the enhancement layer leaves room for a visual improvement. For the typical bit-rates of the HDTV enhancement layer (16 Mbit/s or more), only critical sequences will alow some improvement due to spatial scalability, e.g., in vivid motion or just after a scene cut.

5 Comparison of hardware complexity

We will just compare the hardware necessary for source decoding, i.e., a non-scalable decoder for simulcast and a scalable decoder for hierarchical source coding. The hardware for the hierarchical transmission chain, i.e., the layered modulation and channel coding, is the same in simulcast and in scalable source coding.

Also for the hardware, we make a distinction between the case of quality scalability and of resolution scalability.

5.1 Quality scalability

In quality scalability, Selinger pointed out that an SNR scalable chip requires no additional memory compared to a non-scalable decoder [8]. The extra chip area required for SNR scalability is estimated to be at most 20 %. With a time multiplex of base and enhancement layer data, the extra necessary chip area could be reduced to a few percents. However, chip costs are mainly influenced by the package and not by directly by chip area. The package and pinning is similar in base and enhancement layer.

5.2 Spatial scalability

In this case, the cost of the scalable decoder is higher than the cost of a non-scalable decoder (for decoding the enhancement layer of simulcast). The scalable decoder needs on top of the non-scalable decoder a smaller decoder, including memory, for the base layer. If there should be a fall-back possible to the upconverted base layer in simulcast, then the hardware for upconversion is common to both scalable and non-scalable decoders.

Also in this case, the extra chip area in the scalable decoder could be reduced by a time multiplex of base and enhancement layer. In any case, the principal extra cost in the spatially scalable decoder is for the memory chips of the base layer. Therefore, in a scalable HDTV-decoder with a TV base layer, the decoder cost is approximately 1,3 times higher [8].

6 ADTT Simulations

6.1 Objectives of the experiment

Within the framework of EUREKA ADTT, two main broadcast scenarii had to be compared in order to contribute to the work of DVB on the introduction of digital HDTV: simulcast and embedded. Therefore, a